

探究武汉房价的时空分布差异和影响房价的因素

刘子煜¹ 2019202050035

1、地图学与地理信息系统，武汉大学资源与环境科学学院，武汉大学

摘要：城市住宅价格分布，是规划与环境区位的作用结果。本文通过网络爬虫获取武汉市 2018 年、2019 年部分小区房价信息，通过普通 Kriging 插值方法分析房价的空间时间分异规律，并通过多元线性回归分析研究对武汉市房价分布影响最大的区位因素。研究结果表明江岸区和武昌区为高房价集中区域，且 2019 年相较 2018 年武汉中心城区房价有小幅下降趋势，且新增多个高房价聚集中心。而选取的影响因子与房价间的多元线性回归相关性较小，但其中小区距武汉市 3 个火车站的距离平均值对房价的影响最大。

关键字：房价；影响因子；Kriging 插值；空间分布；回归分析

1 引言

随着城市经济的繁荣发展，住房价格成为居民和投资者长期关注的焦点，尽管国家出台了“限购令”的政策并对房价进行了一些调控，房价上涨依旧是不可避免的趋势，同时城市的再规划与扩张开发，空间重构和城市资源配置的重新分配，改变了房价的空间分布规律，更加凸显了房价时空异质性 [1]。因此，房价历来受到各界学者的广泛关注，但各有侧重，经济方面包括房价高涨的原因 [2]、房价上涨对国民消费 [3]、工业发展 [4]等的影响；以及空间地理方面，城市空间结构、区位环境对房价的影响以及房价的空间分异规律 [5, 6]等。

居住地，即住房，作为人类各项生产生活活动的衔接点，其地理位置无疑受到人们的更多关注，因而房价与地理位置相关的研究显得更加重要。这方面的研究也主要分为两类，一类是研究房价的空间结构：如梅志雄，黎夏 [7]，基于 ESDA 和 Kriging 方法分析了东莞市住宅价格空间结构；王梦晗，刘纪平，王勇 [8]等人用路网距离约束的 GTWR 模型对北京房价进行了分析；张海永，关鹏 [9]利用非对称复合矩阵对房价空间的自相关性进行了研究。另一类是对房价影响因素的研究：如孙雯雯、吕学昌、孔德智 [10]利用普通克里金差值研究了城市规划在房价调控上的作用机制；李静 [11]通过交通可达性分析了上饶市中心城区住房价格；郭贵海 [12]利用特征价格模型基于房价的时空分布规律量化了外部因素对房价的影响程度。

本文利用小区可步性指数，探索可步性和房价之间的关系。另一方面，考虑到各个城市的城市规划、经济结构以及社会文化、社会氛围的差异性，影响房价的各因子之间存在很大的差

异，通过对因子的主成分分析，找出影响武汉市最主要的环境影响因子，并对武汉城市规划提出一定的看法。

2 研究区域与数据

以武汉市为研究区域，选择住宅小区作为基本研究单元，研究武汉市 2018、2019 年的房价的分布规律。依托“链家网”（<https://wh.lianjia.com/xiaoqu/>）房地产交易网站，爬取武汉市 18、19 年小区的房价数据。应用 python 的 bs4 包，通过解析 url 的结构，获得相关信息，包括出售房的小区名、房价、地址、楼层、小区规模、小区物业费。小区的坐标信息通过 urllib 包和正则表达式获取，基于其空间坐标建立小区空间属性数据库。

网上爬取的小区房价数据具有实时性和可靠性。经整理，2018 年的小区房价数据共有 2969 条，其中有效样本中有 2173 个；2019 年小区房价数据共有 5266 条，中有效样本中有 3769 个。元数据坐标基于百度坐标系，为将数据绘制在城市行政区划上，将数据坐标经纬度转换为 WGS1984 地理坐标，并系绘制在地图上，样本数据主要集中在武汉城市中心区，包括洪山区、江岸区、江汉区、硚口区、汉阳区、武昌区、青山区。为集中研究区域，确保研究的准确性，根据数据范围对武汉城区进行裁剪，只保留有样本数据的部分。

图 1 小区价格分布情况图说明，房价较高的地区集中在城中长江附近，主要分布在江岸区、江汉区、武昌区、汉阳区和洪山区，图 2 为武汉各行政区房价平均值及平均房价随时间波动差价分布图。即从 2018 年至 2019 年，在有数据的区域，除了东西湖区和蔡甸区，其他主要城区的平均房价都普遍下降，其中青山区房价下降的最多为 2507 元，洪山、武昌和江汉区的平均房价下降 1000 元以内，而从 2018-2019 年，武昌、江岸和洪山区的平均房价皆遥遥领先，其中武昌和江岸区的平均房价两年来皆保持在 20000 元之上。平均房价最低的蔡甸区房价虽有所增长，但相较于其他区，平均房价仍然较低。

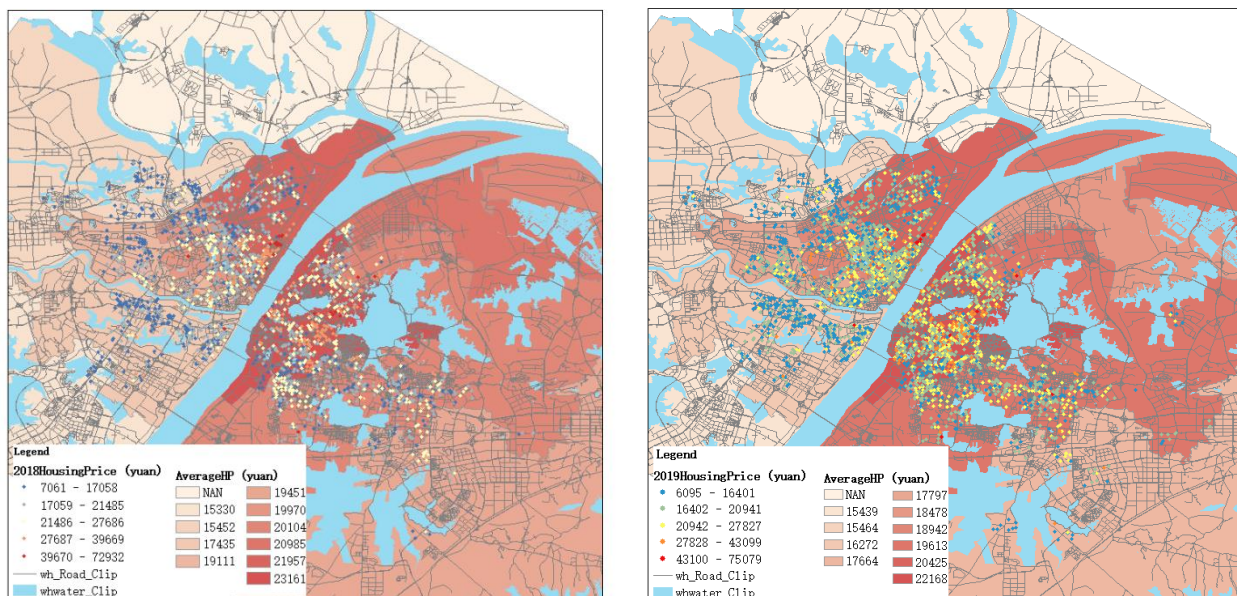


Figure 1 武汉市中心城区小区房价分布图（左 2018 年、右 2019 年）

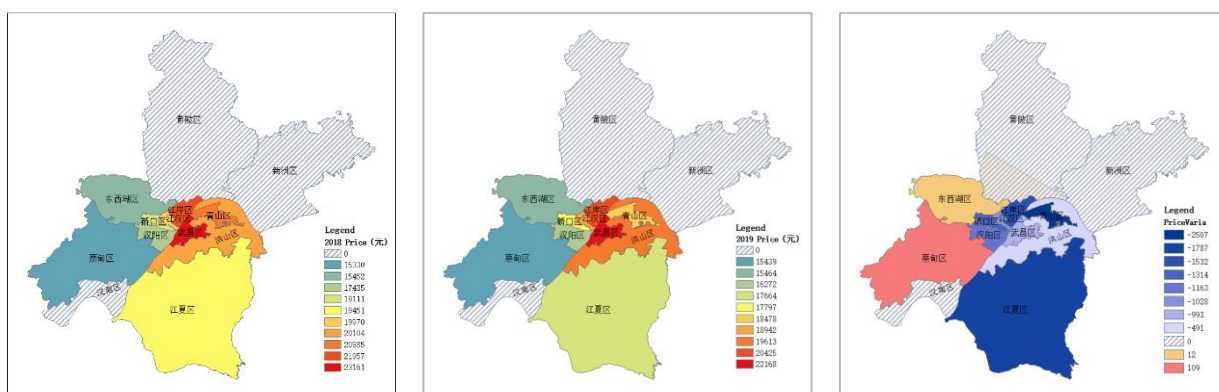


Figure 2 武汉市行政区划平均房价分布情况图（左 2018 年、中 2019 年、右差价）

为探究武汉小区房价的分布时空规律及差异，研究通过克里金差值、等值线分析和趋势面分析对 2018、2019 年的房价分布进行分析，并通过高德 api，从餐饮服务、风景名胜、购物服务、休闲娱乐、金融保险服务、生活服务、科教文化服务、政府机构和医院等多个兴趣点获取 poi 坐标，作为房价的影响因子分析小区区位因素中对小区房价影响最大的影响因素。

3 研究方法

1. 数据检验

首先对样本数据运用 ArcGIS 软件进行统计分析，作直方图 (Histogram) 和正态 QQplot 检验，得到样本数据的统计特征，如表 1：

表格 1 武汉市城区小住宅价格样本数据统计特征

年份	样本数量	最大值	最小值	平均值	标准差	偏度	峰度
2018	2138	72932	7061	20560	6040.923	2.1483	8.90059
2019	3769	75079	6095	19630	5829.099	2.2653	10.18755

由表中两年房价的分布可以看出，房价数据偏度皆约为 2，故房价呈右偏分布，经过 Box-Cox 变换后更符合正态分布（图 3）。

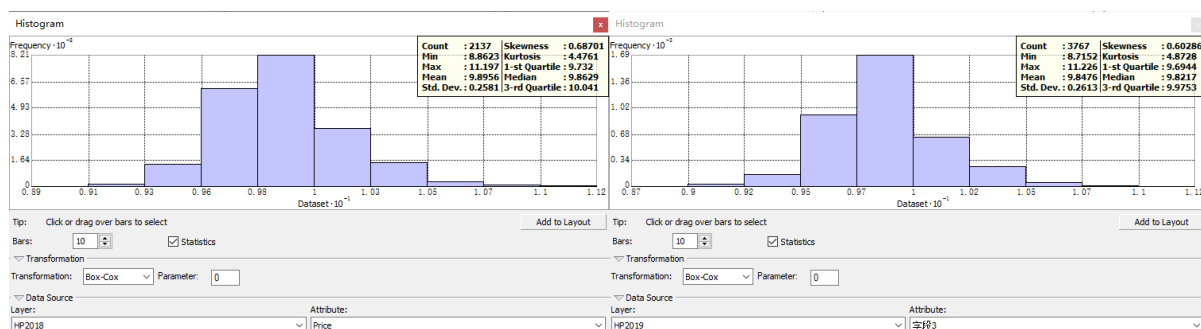


Figure 3 数据变换后直方图（左 2018 年、右 2019 年）

原始的 Normal QQplot 存在偏离正态分布的现象，经过 Box-Cox 变换的 Normal QQplot 接近正态分布。如图 4。

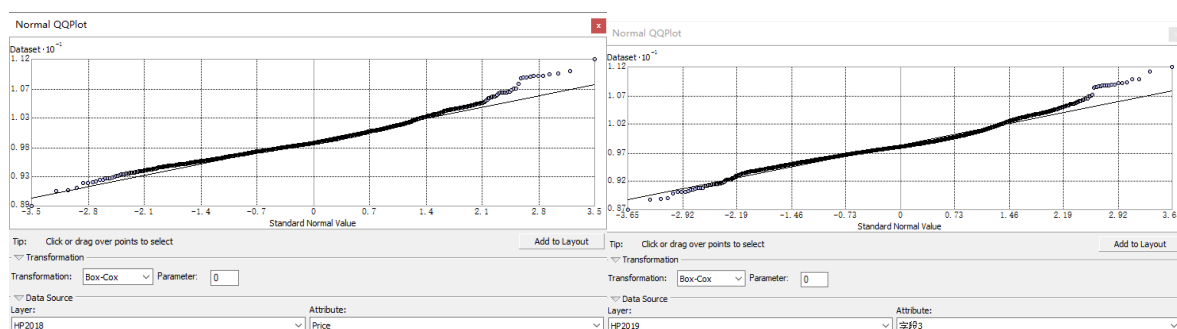


Figure 4 数据变换后的正态 QQplot 图（左 2018 年、右 2019 年）

2. 小区房价的自相关性

影响小区房价的最主要因素是区位，而根据 Toplor 地理学第一定律，空间相近的小区拥有相似的区位环境，故其房价之间也存在一定的相关性。用 Moran's I 分别对两年的小区房价空间分布的自相关性进行测度，判断样本数据及的集散程度。

利用 ArcGIS 软件，以房价为属性，对小区空间数据集建立以距离权重的全局一元 Moran's I，并用兰指数检验方法对小区房价的空间分布特征进行分析。图 5 为 2018、2019 两年武汉小区房价的基于莫兰指数的空间自相关分析。

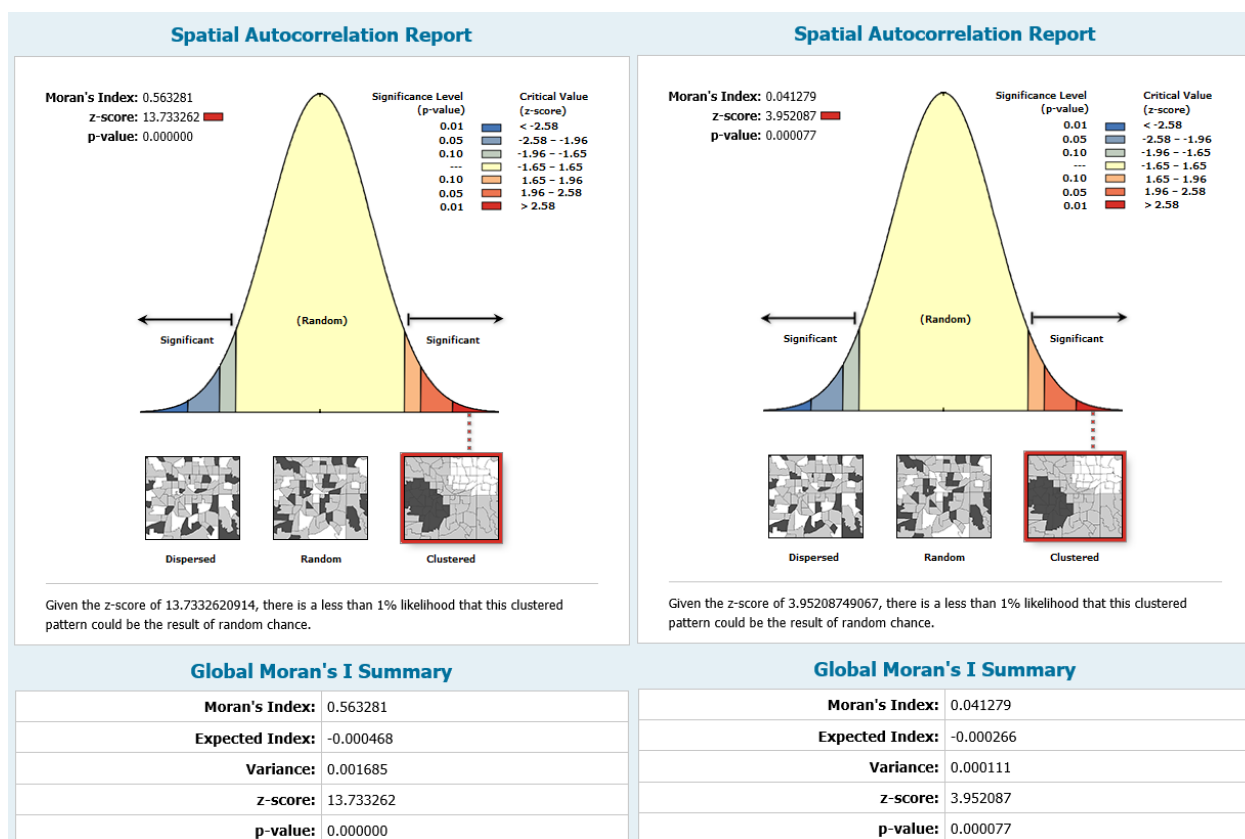


Figure 5 小区房价基于莫兰指数的空间自相关分析报告（左 2018 年、右 2019 年）

Moran's I 值介于-1 和 1 之间， $I > 0$ 表示空间正相关，空间实体成聚合分布。从两年的房价空间分布报告可以看出，武汉市的房价一直呈现空间聚集的分布状态，且 Z 值皆为正并大于 2.58，p 值皆接近 0.00，数据结果分布在正态分布的右端，为聚集型，且结果具有可信度。

3. 小区房价 Kriging 插值和等值线分析

Kriging 插值法是以空间自相关性为基础，利用半变异函数的结构性，对有限区域内的区域化变量取值进行无偏最优估计的一种方法，它不仅考虑了距离关系，而且通过变异函数和结构分析，考虑了已知样本点的空间分布及与未知样本点的空间方位关系，还利用了已有观测值空间分布的结构特征，使其估计结果比传统的方法更为精确，更有效地避免了系统误差的出现 [13]。

故确定变异函数是 Kriging 插值方法的核心问题。对于任意待估点的估计值均可以通过待估测点范围内的 n 个观测样本的线性组合得到，点 Kriging 内插法估计任一点 X_0 处的房价的算法如下：

$$Z(X_0) = \sum_{i=1}^n \lambda_i Z(X_i)$$

式中， $Z(X_i)$ 是 X_i 点处的房价； λ_i 为权重系数，是各已知房价 $Z(X_i)$ 的在估计 $Z(X_0)$ 时影响大小的系数，是在变异函数基础上，将半方差带入 Kriging 方程组求解得到的，其和等于 1 [7]。

根据以上方法，采用普通 Kriging 法对样本点进行插值，得到的武汉市 18、19 两年小区价格空间分布插值图，如图 6。考虑到样本主要分布在武汉中心城区，且两年间的小区分布范围并不相同，插值区域存在差异，但可以发现 2019 年武汉城区的房价较高的地区范围逐渐缩小，但明显形成了以黄家湖军运村为中心的高房价区域。

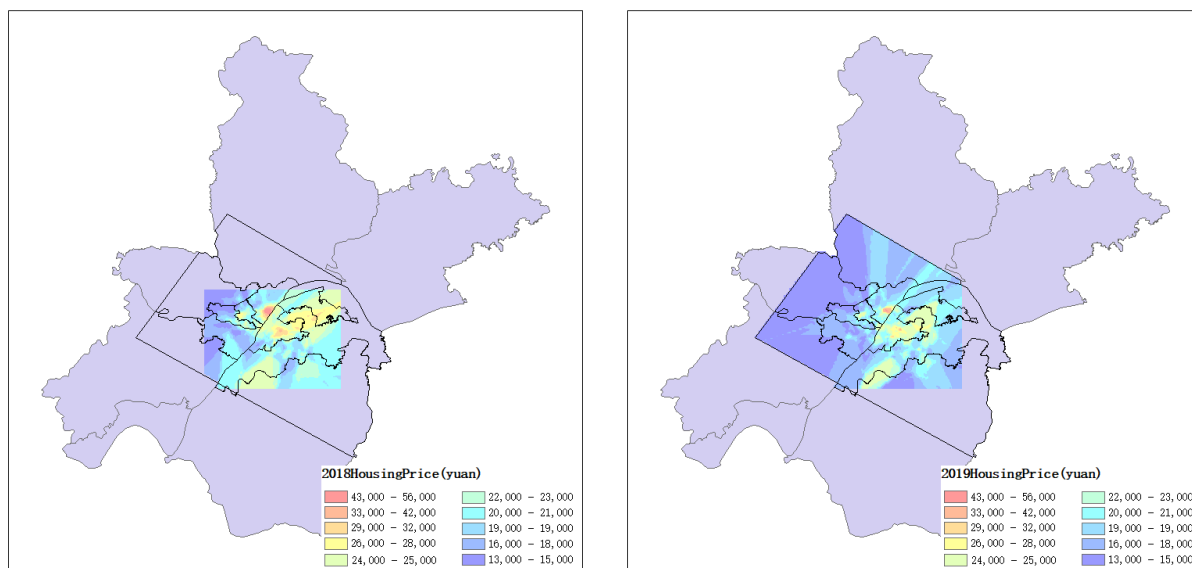


Figure 6 武汉小区房价空间分布的 Kriging 插值（左 2018 年、右 2019 年）

以 2000 元/平方米为间隔，对中心城区的房价绘制等值线。如图 7。从图中可知，江岸区江滩附近为高房价聚集中心，且房价在该区域呈现陡增状态，直至距江滩最近达到峰值。但 2019 年武汉中心城区的房价普遍下降，且高房价中心范围皆有缩小的趋势，但出现多中心的高房价分布趋势，而主要的武昌、江岸区的高房价中心变化较小。

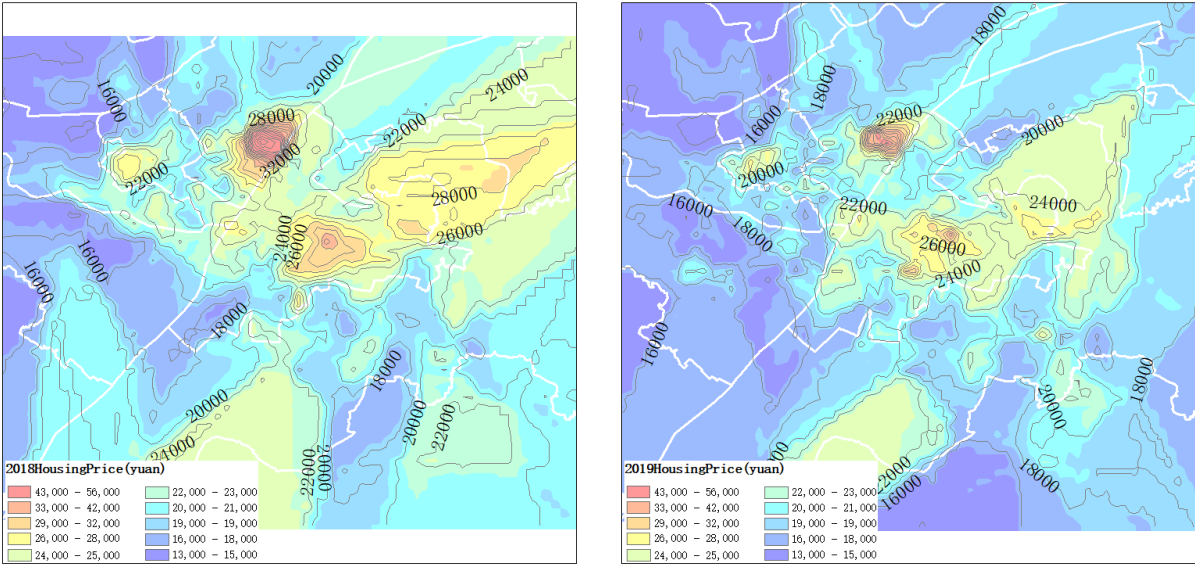


Figure 7 武汉中心城区房价等值线分布（左 2018 年、右 2019 年）

4. 影响因子分析

影响房价的因素有很多，不同的区位环境有不同的主导因素，为探讨影响武汉市房价分布的主要因素，我们采用 2018 年的房价数据，从各个服务设施类中选取影响房价的区位因素，并通过多元回归分析，找出最有可能影响房价的因子。

1) 确定影响因子

从社区属性、区位交通、周边配套设​​施和其他增值属性对影响社区房价空间分异特征变量进行描述，如表 1：

Table 1 房价影响因子说明表

特征分类	解释变量	变量代称	变量描述	备注
社区属性	居住规模		住宅数	住宅总套数
	物业费标准		每月每平米	
区位交通	中心区位			
	环境区位	n_w	到水域最短路径	临江、滨湖
	内部交通	Post_w	到最近两个地铁站出入口最短路径之和	公共交通网点
	路网密度	Cross_D	2.4km 内交叉口密度	

周边配套	生活配套	Res_w	大型超市、便利店	可步性衡量
	商业配套	Smal_w	商场	可步性衡量
	商务配套	Bank_w	银行	可步性衡量
	休闲配套	SMarket_w ; Sports_w; Ent_w	公园、体育场馆、娱乐等	可步性衡量
	医疗保健	Hos_w	医院	可步性衡量
增值属性	规划属性	Under	是否位于规划新城区 范围内	
	学区属性	Pris_w	学校附近	可步性衡量
	景观稀缺	MinD_Inter	风景名胜	
	房源稀缺	com_relati	1 km 范围住宅价值总量与住房数量的关系	各小区住宅均价/住房总套数

2) 多元线性回归

多元线性回归模型（Multivariable Linear Regression Model），简称 MLR。一般假设解释变量和房价之间存在线性关系，选用多个因子来解释房价的空间变化。

$$y = a_0 + \sum_{i=1}^n a_i x_i + \varepsilon \quad i = 1, 2, \dots, n$$

式中， y 是房价， a_0 是常数， a_i 是解释变量 x_i 的系数， n 表示解释变量的个数， x_i 则表示第 i 个影响因子。

4 实验结果

Table 2 多元线性回归模型摘要

模型摘要

模型	R	R 方	调整后 R 方	标准估算的误差	更改统计				
					R 方变化量	F 变化量	自由度 1	自由度 2	显著性 F 变化量
1	.379 ^a	.143	.136	5616.019	.143	19.700	18	2118	.000

a. 预测变量：(常量), com_relati, Res_w, n_w, Bus_w, MinD_Inter, Train_D_A, park_w, Cross_D, D_College_, Under, Post_w, Pris_w, Smal_w, Ent_w, Sports_w, Hos_w, SMarket_w, Bank_w

表 2 所示为 SPSS 对 19 个因子进行多元线性回归构建的模型，模型 R² 为 0.379，因变量与解释变量之间存在微弱的线性关系，但不明显，显著性检验为 0。

Table 3 多元线性回归系数表

		系数 ^a						
模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间	
		B	标准误差	Beta			下限	上限
1	(常量)	31686.439	979.231		32.359	.000	29766.085	33606.793
	Bank_w	-29.824	28.689	-.055	-1.040	.299	-86.085	26.437
	Bus_w	-102.441	60.304	-.038	-1.699	.090	-220.702	15.820
	park_w	59.254	29.684	.054	1.996	.046	1.041	117.467
	Hos_w	-25.734	24.188	-.049	-1.064	.287	-73.170	21.701
	Post_w	-535.829	151.147	-.113	-3.545	.000	-832.240	-239.417
	Pris_w	331.364	123.285	.102	2.688	.007	89.593	573.136
	Res_w	90.698	45.239	.115	2.005	.045	1.980	179.416
	Smal_w	-135.069	39.562	-.124	-3.414	.001	-212.654	-57.485
	Sports_w	305.243	61.316	.211	4.978	.000	184.997	425.490
	SMarket_w	-119.563	23.469	-.239	-5.095	.000	-165.588	-73.538
	Ent_w	-4.956	30.856	-.006	-.161	.872	-65.467	55.554
	Under	73.627	108.522	.021	.678	.498	-139.195	286.448
	Cross_D	-22.996	9.381	-.067	-2.451	.014	-41.392	-4.600
	Train_D_A	-.806	.065	-.329	-12.389	.000	-.933	-.678
	n_w	-.191	.275	-.015	-.693	.489	-.731	.349
	D_College_	-.030	.132	-.006	-.229	.819	-.289	.228
	MinD_Inter	-.509	.185	-.074	-2.752	.006	-.871	-.146
	com_relati	.016	.107	.003	.150	.881	-.194	.226

a. 因变量：Price

表 3 为多元回归的系数，由该表系数和 t 检验归纳可知大型超市的可步性、武汉 3 个火车站到小区的平均距离、运动场所的可步性，商场的可步性及附近小学的可步性对房价的影响较大。已知到火车站的平均距离与房价呈负相关，即意味着平均距离越短，房价越高，而考虑武汉市三个火车站的分布，越靠近武汉市中心城区距离火车站越短，符合房价的分布规律。武汉作为中国内陆中心城市，在人员流动方面起着至关重要的作用，大量的人流吞吐也合理解释了与火车站间的距离对武汉房价的重要影响。

5 结论

武汉市的房价呈星状多中心分布，在城区中心地带房价普遍较高，远离市中心地区房价较低，有较高的空间自相关性，呈现出较高的空间聚集。武汉市小区高房价主要集中于江岸区

和武昌区, 2019 年武汉市房价有普遍下降, 但出现更多的高房价聚集区, 可见城市建设范围的扩大和地铁的逐渐开通带动了附近楼盘的发展。将一些区位环境因素包括增值环境, 如靠近名胜古迹, 结合可步性指标中的小指标对小区房价进行多元线性分析, 结果显示相关性较小。从模型系数和置信度观察发现, 小区到火车站的距离对房价影响较大。

本研究存在很多问题, 包括实验数据的处理方法有待改进, 数据精度有待提升。同时回归分析的方法最需要改善。包括解释因子类别和数值的确定。研究同时用了地理加权回归对文中设定的因子进行了回归分析, 效果依然不理想。

参考文献

1. 梅志雄 and 黎夏, 基于地统计分析的房价空间分布差异研究——以东莞市 2006 年普通住宅价格为例. 华南师范大学学报(自然科学版), 2007(04): p. 120-126.
2. 王弟海, 管文杰, and 赵占波, 土地和住房供给对房价变动和经济增长的影响——兼论我国房价居高不下持续上涨的原因. 金融研究, 2015(01): p. 50-67.
3. 颜色 and 朱国钟, “房奴效应”还是“财富效应”?——房价上涨对国民消费影响的一个理论分析. 管理世界, 2013(03): p. 34-47.
4. 王文春 and 荣昭, 房价上涨对工业企业创新的抑制影响研究. 经济学(季刊), 2014. 13(02): p. 465-490.
5. 曾晖, 城市住宅价格时空分布规律研究. 2012, 南京林业大学.
6. 赵自胜, 城市商品住宅价格空间分异研究. 2010, 河南大学.
7. 梅志雄 and 黎夏, 基于 ESDA 和 Kriging 方法的东莞市住宅价格空间结构. 经济地理, 2008(05): p. 862-866.
8. 王梦晗, et al., 路网距离约束的 GTWR 模型应用——以北京市房价为例. 测绘科学, 2018. 43(04): p. 133-137.
9. 张海永 and 关鹏, 基于非对称复合矩阵的房价空间自相关性研究. 巢湖学院学报, 2017. 19(03): p. 1-9.
10. 孙雯雯, 吕学昌, and 孔德智, 城市规划在房价调控上的作用机制研究——基于 GIS 以济南市中心城区为例的实证分析. 现代城市研究, 2014. 29(12): p. 39-45.
11. 李静, 基于交通可达性的上饶市中心城区住房价格模拟分析. 2017, 江西师范大学.
12. 郭贵海, 城市住宅价格演变规律空间影响因子相关性研究. 2012, 中国矿业大学(北京).
13. 蒋芳 and 朱道林, 基于 GIS 的地价空间分布规律研究——以北京市住宅地价为例. 经济地理, 2005(02): p. 199-202.

